Measurement Invariance in Noncognitive Measures: Validity Approach Using Explanatory Item Response Modeling

Jose Palma[1], Youngsoon Kang[1], Okan Bulut[2], Julio Cabrera[1]

[1]University of Minnesota
[2]University of Alberta

Minnesota Youth Development Research Group
www.mnydrg.com

April 2018

Paper presented at the annual meeting of the
National Council on Measurement in Education, New York, NY.

Citation:

Palma, J., Kang, Y., Bulut, O., & Cabrera, J. (2018, April). *Measurement invariance in noncognitive measures: Validity approach using explanatory item response modeling.* Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

Measurement Invariance in Noncognitive Measures: Validity Approach Using Explanatory Item Response Modeling

Abstract

Using a partial-credit explanatory item response model, we examine validation concerns of score interpretation of two developmental measures, Social Competence and Empowerment across grade levels. We find significant interactions between items and grade levels for both constructs implying different perceptions of items across grades and potentially influencing score interpretation.

Introduction

As the role of schools and districts continue to change to include the development of 21st century skills (social and emotional learning) as part of their curriculums (Griffin, McGaw, & Care, 2012), educators, policymakers, and youth development researchers have focused their attention to the roles of noncognitive factors (developmental assets or social-emotional skills) in learning and achievement. The importance of developmental assets in learning and achievement has been widely studied (e.g., Scales, Benson, & Mannes, 2006). However, measurement of developmental assets is relatively new and the psychometric work in this area is minimal. There is a need to incorporate the expectations from the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) for the assessment of noncognitive factors.

Given the importance of noncognitive factors and the increased efforts to measure them, validity evidence must be provided to support the uses and interpretations of scores drawn from these measures (Kane, 2013). Measurement invariance is to create a common interpretation framework and provide empirical support for interpretations across test-takers (Millsap, 2011). To go along with the NCME theme, *making assessment a stronger force for positive impact on teaching and learning*, validity evidence is critical to provide sound and fair inferences from scores across test-takers when using non-cognitive measures.

When using measures of social-emotional learning (SEL) across grade levels, the nature of the construct and the diversity of the population must be taken into account

(Castellano & Kolen, 2016; Knight & Zerr, 2010). Youth go through developmental changes that are biological, cognitive, and social-emotional (Santrock, 2011, p. 29) that could potentially shift the construct being assessed (Bulut, Palma, Rodriguez, & Stake, 2015; Castellano & Kolen, 2016). However, measurement invariance analyzes are rarely applied to noncognitive measures.

We use two measures of social-emotional learning from the Developmental Asset Profile (DAP; Search Institute, 2013), *Empowerment* and *Social Competence*, to examine item characteristics are influenced by youth's developmental changes across grade levels. The purpose of the study is to address concerns of measurement invariance and examine drift (shift) in the response category thresholds of the items across grade levels.

Research Question

Are the item threshold parameters for *Empowerment* and *Social Competence* consistent across grade levels? We hypothesize, consistent with developmental theory, that if the students from different grade levels perceive the constructs of *Empowerment* and *Social Competence* similarly, the item threshold parameters should be very similar for all students, regardless of the grade levels. If, however, there are major differences in the way students from different grade levels perceive these constructs (i.e., construct shift), then there should be significant differences in the item threshold parameters.

Methods

Data

In this study, we used data from the 2016 Minnesota Student Survey (MSS; MN Department of Education, 2017). The MSS is designed by an interagency team from the MN Departments of Education, Health, Human Services, and Public Safety. The survey is administered every three years to students in grades 5, 8, 9, and 11. Approximately 85% of the MN public school districts participated in the 2016 MSS administration and the sample closely resembles the state population in terms race, ethnicity, participation in special education, and free and reduced-price lunch. We used a sample of 168,733 students (Female = 50.3%) in grades five (n=41,865), eight (n=44,983), nine (n=45,309), and eleven (n=36,576). Table 1 provides a summary of the sample characteristics by grade.

Table 1

Participating Sample by Race and Grade

Race/Ethnicity	*Grade 5*	*Grade 8*	*Grade 9*	*Grade 11*	*Total*
American Indian	2577	2517	2255	1313	8662
Asian only (not Hmong) or Pacific Islanders	1428	1609	1623	1250	5910
Black only (not Somali)	2781	2408	2065	1552	8806
White	26442	29393	31121	26357	113313
Multiple Race or Ethnicity	1200	1785	1612	1164	5761
Latino	3902	4838	4213	2989	15942
Somali	1182	958	835	580	3555
Hmong	1122	1131	1261	1130	4644
Total	41865	44983	45309	36576	168733

The MSS includes two measures of developmental skills and supports from the Developmental Asset Profile, *Empowerment* and *Social Competence* (Search Institute, 2013). *Empowerment* is measured with six items (see Table 2). The first four items are found in a 4-point rating scale (strongly disagree, disagree, agree, strongly agree), and the last 2 items are also found in a 4-point rating scale (rarely, sometimes, often, almost always). *Social Competence* is measured with eight items (see Table 2), each item including a 4-point rating scale (rarely, sometimes, often, almost always).

Each measure was submitted to confirmatory factor analysis (CFA) with Mplus (v.7; Muthen & Muthen, 2012). CFA showed that the two scales are unidimensional and have adequate structures. Fit indices, including the comparative fix index (CFI), Tucker-Lewis Index (TLI), and the root mean-squared error of approximation (RMSEA) were used to examine the adequacy of each model structure (Brown, 2015). For *Empowerment*, the fit indices were CFI = .91, TLI = .85, and RMSEA = .23. For *Social Competence*, the fit indices were CFI = .92, TLI = .91, and RMSEA = .08. Subsequently, both measures were scaled using the Rasch model with Winsteps (v. 3.92; Linacre, 2016).

Table 2

Items in the Measures of Empowerment and Social Competence

Empowerment	Social Competence
a. Feel safe at school	a. Resist dangerous/unhealthy things
b. Feel safe in neighborhood	b. Build friendships
c. Feel safe at home	c. Express feelings in proper ways
d. Feel valued by others	d. Plan ahead and make good choices
e. Included in family tasks and decisions	e. Avoid bad influences
f. Given useful roles and responsibilities	f. Resolve conflicts without violence
	g. Accept people who are different
	h. Sensitive to others' needs/feelings

Analytical Model

The Partial Credit Model (PCM) is an item response theory model for polytomously scored items with ordered response categories. PCM can also be expressed as a generalized linear mixed model, and thus, it can be extended to an explanatory form by including item and person related covariates in the model. For polytomous item i $(i = 1, 2, 3, \dots, I)$ with adjacent response categories indexed by j $(j = 1, 2, 3, \dots, J)$, the log-odds of selecting response category j over $j - 1$ on for person n can be written as:

$$\log\left(\frac{P_{nij}}{P_{ni(j-1)}}\right) = \theta_n - (\delta_i + \tau_{ij}), \quad (1)$$

where θ_n represents the latent trait of person n with a normal distribution $N(\mu_n, \sigma_n^2)$. Although δ_i is often considered the overall item difficulty, this parameter represents the location of the threshold between the first ($j = 0$) and second ($j = 1$) response categories for item i. τ_{ij} represents the distance between the other thresholds. For example, if item i has three response categories, τ_{ij} would represent the distance between the $(j - 2)/(j - 1)$ threshold and the $(j - 1)/j$ threshold.

In the explanatory partial-credit model (PCM), the log-odds of selecting response j over $j - 1$ on item i for person n can be written as:

$$\log\left(\frac{P_{nij}}{P_{ni(j-1)}}\right) = \mathbf{Z_{nij}}\theta_n - \mathbf{X'_{nij}}(\delta_i + \tau_{ij}), \quad (2)$$

where $\mathbf{Z}_{nij}$ is a matrix of fixed- and random-effects related to the latent trait θ_n distributed as $N(\mu_n, \sigma_n^2)$. $\mathbf{X}_{nij}$ is a matrix of fixed- and random-effects related to individual items. $\boldsymbol{\delta_i}$ is the location of the threshold between the first ($j = 0$) and second ($j = 1$) response categories for item i. τ_{ij} is the distance between the other thresholds. The models are estimated with *lme4* (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2017).

In this study, we used the 2016 MSS data to investigate drift (or shift) in the responses categories on the item responses in the *Empowerment* and *Social Competence* measures. Three models were evaluated for each measure:

1. Model 1: Partial credit model + item predictors
2. Model 2: Partial credit model + item predictors + person predictors (grade levels)
3. Model 3: Partial credit model + predictors + their interaction (item x grade levels)

With the three models, we can estimate model fit to the item response data conditioning on item characteristics (Model 1), and compare the fit to the models where we can account for item and person characteristics (Model 2), and their interaction (Model 3). Model 3 is of interest since it includes the item by grade interaction. Model fit indices are evaluated, as well as model parameters.

Results

For each scale, three models were fit to examine the functioning of the ordinal responses from the rating scale items, given person characteristics (grade level). The three models include, the PCM model with main effects for items, the PCM model with main effects for items and person characteristics, and the full PCM model with main effects and interaction between the item and person characteristics (grade level). Results are discussed here briefly and a more complete output of the *lme4* results is provided in Appendices A-C (*Social Competence*) and Appendices D-F (*Empowerment*).

Table 3 displays the fit indices of the three models for *Empowerment* and *Social Competence.* By examining AIC and BIC results, the PCM model with main effects for

both item and person characteristics fits better than Model 1 and Model 3. This is the case for both constructs, *Empowerment* and *Social Competence.*

Table 3

Model Fit Indices for the Explanatory Partial Credit Models

Model	AIC	BIC	logLikelihood	Deviance
Social Competence				
Model 1	2435081	2435394	-1217516	2435031
Model 2	2429436	2429786	-1214690	2429380
Model 3	2444500	2444812	-1222225	2444450
Empowerment				
Model 1	1644225	1644457	-822094	1644188
Model 2	1640956	1641224	-820546	1640912
Model 3	1648521	1648752	-824241	1648483

Note. Model 1: main effects for items. Model 2: main effects for items and persons. Model 3: includes interactions for item and person characteristics.

The grade level effect is significant ($p < .01$) in the two models where this variable is included for the two measures of *Empowerment* and *Social Competence.* The interaction effects (item x person characteristics; Model 3) are significant ($p < .01$) for both measures. The interaction effect indicates that students from different grade levels perceive the constructs of *Empowerment* and *Social Competence* differently. However, if the students from different grade levels perceive the constructs similarly, the item threshold parameters should be very similar for all students, regardless of the grade levels. If, however, there are major differences in the way students from different grade levels perceive these constructs (i.e., construct shift), then there should be significant differences in the item threshold parameters.

Summary results of Model 3 for each construct are provided in Table 4. As a general trend, the second threshold distance is smaller than the third threshold distance across items in both constructs. A negative interaction coefficient indicates that it takes less of the trait to be in that response category for later grades than earlier grades.

Table 4

Interaction Explanatory Partial Credit Model for Social Competence and *Empowerment*

	Social Competence		Empowerment	
	Coefficient	SE	Coefficient	SE
Main Effects				
Item a x Grade	0.23	0.002	0.37	0.003
Item b x Grade	0.28	0.002	0.39	0.004
Item c x Grade	0.23	0.001	0.41	0.005
Item d x Grade	0.31	0.002	0.28	0.001
Item e x Grade	0.26	0.002	0.31	0.002
Item f x Grade	0.31	0.002	0.35	0.002
Item g x Grade	0.33	0.003		
Item h x Grade	0.26	0.002		
Interaction Effects				
Item a x Grade x Threshold 2	-0.12	0.002	-0.01*	0.003
Item b x Grade x Threshold 2	-0.15	0.002	0.04	0.004
Item c x Grade x Threshold 2	-0.17	0.002	0.07	0.005
Item d x Grade x Threshold 2	-0.21	0.001	-0.18	0.001
Item e x Grade x Threshold 2	-0.16	0.002	-0.15	0.002
Item f x Grade x Threshold 2	-0.19	0.002	-0.17	0.002
Item g x Grade x Threshold 2	-0.09	0.003		
Item h x Grade x Threshold 2	-0.13	0.002		
Item a x Grade x Threshold 3	-0.17	0.002	-0.35	0.003
Item b x Grade x Threshold 3	-0.32	0.002	-0.31	0.004
Item c x Grade x Threshold 3	-0.38	0.002	-0.25	0.005
Item d x Grade x Threshold 3	-0.41	0.002	-0.38	0.002
Item e x Grade x Threshold 3	-0.25	0.002	-0.36	0.002
Item f x Grade x Threshold 3	-0.37	0.002	-0.40	0.002
Item g x Grade x Threshold 3	-0.28	0.003		
Item h x Grade x Threshold 3	-0.29	0.002		

Note. All significant at $p < .05$ with the exception of (*).

It has been recognized the difficulty of assessing social-emotional skills across different developmental stages given the nature of the constructs (Griffin et al., 2012; Knight & Zerr, 2010). It is critical to provide sound evidence to support score uses and interpretations from non-cognitive measures for the population being assessed (Kane, 2013). Findings such as this put the psychometric researcher at a crossroads between goals of finding a common ground construct interpretation across grade levels and goals of construct representation (Castellano & Kolen, 2016).

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Brown, T. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Bulut, O., Palma, J., Rodriguez, M. C., & Stanke, L. (2015). Evaluating measurement invariance in the measurement of developmental assets in Latino English language groups across developmental stages. *SAGE Open, 5*(2), 1-18.

Castellano, K. E., & Kolen, M. J. (2016). Comparing test scores across grade levels. In Dorans , N. J., & Cook, L. L. (Eds.). *Fairness in educational assessment and measurement.* New York: Routledge.

Griffin, P., McGaw, B., & Care, E (2012). *Assessment and teaching of 21st century skills.* New York: Springer.

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50,* 1-73.

Knight, G. P., & Zerr, A. A. (2010). Informed theory and measurement equivalence in child development research. *Child Development Perspectives, 4*(1), 25-30.

Linacre, J.M. (2016). Winsteps® (Version 3.92.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved from http://www.winsteps.com/

Millsap, R. E. (2011). *Statistical approaches to measurement invariance.* New York: Routledge.

MN Department of Education (2017). *Minnesota Student Survey.* Roseville, MN: Author. Retrieved from https://education.mn.gov/MDE/dse/health/mss/

Muthén, L.K., & Muthén, B.O. (2012). Mplus. (Version 7). [Software program]. Los Angeles, CA: Authors.

R Core Team. (2017). *lme4*. Retrieved from https://CRAN.R-project.org/package=lme4

Santrock, J. W. (2011). *Educational psychology.* New York: McGraw-Hill.

Search Institute. (2013). *Developmental Assets Profile: Technical summary*. Minneapolis, MN: Author. Retrieved at http://www.search-institute.org/surveys/dap

Scales, P.C., Benson, P.L., & Mannes, M. (2006). The contribution to adolescent well-being made by nonfamily adults: An examination of developmental assets as contexts and processes. *Journal of Community Psychology, 34*(4), 401-413.

Appendix A

Explanatory Unconditional Partial Credit Model of the 8-item Measure of Social Competence

```
> summary(model.epcm2)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: PCM ~ -1 + item + item:PCMcategory + (1 | studentnumber)
   Data: soccomp3
Control: control

     AIC      BIC   logLik deviance df.resid
 2435081  2435394 -1217516  2435031  1983561

Scaled residuals:
    Min      1Q  Median      3Q     Max
-7.8544 -0.8117  0.4280  0.6880  7.0276

Random effects:
 Groups        Name        Variance Std.Dev.
 studentnumber (Intercept) 1.269    1.126
Number of obs: 1983586, groups:  studentnumber, 162024

Fixed effects:
                           Estimate Std. Error z value Pr(>|z|)
itemY60c                    1.91838    0.01279  150.02   <2e-16 ***
itemY60d                    2.37681    0.01356  175.24   <2e-16 ***
itemY60e                    1.99494    0.01045  190.96   <2e-16 ***
itemY60i                    2.72297    0.01372  198.51   <2e-16 ***
itemY60j                    2.30244    0.01325  173.73   <2e-16 ***
itemY60k                    2.58320    0.01374  187.97   <2e-16 ***
itemY60m                    2.82669    0.02232  126.66   <2e-16 ***
itemY60q                    2.21183    0.01324  167.01   <2e-16 ***
itemY60c:PCMcategorycat_3 -0.88387    0.01492  -59.24   <2e-16 ***
itemY60d:PCMcategorycat_3 -1.28852    0.01505  -85.61   <2e-16 ***
itemY60e:PCMcategorycat_3 -1.41978    0.01194 -118.93   <2e-16 ***
itemY60i:PCMcategorycat_3 -1.81452    0.01496 -121.29   <2e-16 ***
itemY60j:PCMcategorycat_3 -1.29567    0.01501  -86.34   <2e-16 ***
itemY60k:PCMcategorycat_3 -1.51639    0.01515 -100.11   <2e-16 ***
itemY60m:PCMcategorycat_3 -0.71310    0.02429  -29.35   <2e-16 ***
itemY60q:PCMcategorycat_3 -1.07247    0.01489  -72.04   <2e-16 ***
itemY60c:PCMcategorycat_4 -1.29116    0.01423  -90.76   <2e-16 ***
itemY60d:PCMcategorycat_4 -2.70790    0.01485 -182.29   <2e-16 ***
itemY60e:PCMcategorycat_4 -3.22838    0.01270 -254.28   <2e-16 ***
itemY60i:PCMcategorycat_4 -3.51278    0.01514 -232.07   <2e-16 ***
itemY60j:PCMcategorycat_4 -2.07794    0.01458 -142.55   <2e-16 ***
itemY60k:PCMcategorycat_4 -3.06067    0.01504 -203.44   <2e-16 ***
itemY60m:PCMcategorycat_4 -2.27918    0.02300  -99.08   <2e-16 ***
itemY60q:PCMcategorycat_4 -2.43500    0.01455 -167.30   <2e-16 ***
---
```

Appendix B

Explanatory Main Effects Partial Credit Model of the 8-item Measure of Social Competence

```
> summary(model.epcm1)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: PCM ~ -1 + item + item:PCMcategory + PCMcategory:grade + (1 |
studentnumber)
   Data: soccomp3
Control: control

     AIC      BIC   logLik deviance df.resid
 2429436  2429786 -1214690  2429380  1983558

Scaled residuals:
    Min      1Q  Median      3Q     Max
-9.6248 -0.8120  0.4223  0.6876  7.4620

Random effects:
 Groups        Name        Variance Std.Dev.
 studentnumber (Intercept) 1.26     1.122
Number of obs: 1983586, groups:  studentnumber, 162024

Fixed effects:
                            Estimate Std. Error z value Pr(>|z|)
itemY60c                    1.191635   0.025774   46.23  < 2e-16 ***
itemY60d                    1.655262   0.026002   63.66  < 2e-16 ***
itemY60e                    1.275947   0.024536   52.00  < 2e-16 ***
itemY60i                    1.995415   0.026163   76.27  < 2e-16 ***
itemY60j                    1.556552   0.026271   59.25  < 2e-16 ***
itemY60k                    1.862738   0.026109   71.34  < 2e-16 ***
itemY60m                    2.106330   0.031438   67.00  < 2e-16 ***
itemY60q                    1.510004   0.025578   59.03  < 2e-16 ***
itemY60c:PCMcategorycat_3   0.143454   0.026526    5.41 6.37e-08 ***
itemY60d:PCMcategorycat_3  -0.267168   0.026312  -10.15  < 2e-16 ***
itemY60e:PCMcategorycat_3  -0.400982   0.024626  -16.28  < 2e-16 ***
itemY60i:PCMcategorycat_3  -0.786968   0.026301  -29.92  < 2e-16 ***
itemY60j:PCMcategorycat_3  -0.250166   0.026741   -9.36  < 2e-16 ***
itemY60k:PCMcategorycat_3  -0.495702   0.026392  -18.78  < 2e-16 ***
itemY60m:PCMcategorycat_3   0.304925   0.032510    9.38  < 2e-16 ***
itemY60q:PCMcategorycat_3  -0.070001   0.025944   -2.70  0.00697 **
itemY60c:PCMcategorycat_4   0.226189   0.026461    8.55  < 2e-16 ***
itemY60d:PCMcategorycat_4  -1.198330   0.026610  -45.03  < 2e-16 ***
itemY60e:PCMcategorycat_4  -1.723970   0.025455  -67.73  < 2e-16 ***
itemY60i:PCMcategorycat_4  -1.999638   0.026809  -74.59  < 2e-16 ***
itemY60j:PCMcategorycat_4  -0.548778   0.026850  -20.44  < 2e-16 ***
itemY60k:PCMcategorycat_4  -1.552804   0.026725  -58.10  < 2e-16 ***
itemY60m:PCMcategorycat_4  -0.767356   0.031912  -24.05  < 2e-16 ***
itemY60q:PCMcategorycat_4  -0.939540   0.026233  -35.82  < 2e-16 ***
PCMcategorycat_2:grade      0.086931   0.002733   31.81  < 2e-16 ***
PCMcategorycat_3:grade     -0.036954   0.001904  -19.41  < 2e-16 ***
PCMcategorycat_4:grade     -0.096921   0.001735  -55.88  < 2e-16 ***
```

Appendix C

Explanatory Full with Interactions Partial Credit Model of the 8-item Measure of Social Competence

```
> summary(model.epcm3)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: PCM ~ -1 + item:grade + item:PCMcategory:grade + (1 | studentnumber)
   Data: soccomp3
Control: control

     AIC      BIC   logLik deviance df.resid
 2444500  2444812 -1222225  2444450  1983561

Scaled residuals:
     Min       1Q   Median       3Q      Max
-12.7771  -0.8242   0.4347   0.6992   9.3587

Random effects:
 Groups        Name        Variance Std.Dev.
 studentnumber (Intercept) 1.284    1.133
Number of obs: 1983586, groups:  studentnumber, 162024

Fixed effects:
                                 Estimate Std. Error z value Pr(>|z|)
itemY60c:grade                   0.235557   0.001546  152.38   <2e-16 ***
itemY60d:grade                   0.278059   0.001639  169.69   <2e-16 ***
itemY60e:grade                   0.230899   0.001248  185.00   <2e-16 ***
itemY60i:grade                   0.314495   0.001648  190.80   <2e-16 ***
itemY60j:grade                   0.263118   0.001554  169.28   <2e-16 ***
itemY60k:grade                   0.305438   0.001679  181.87   <2e-16 ***
itemY60m:grade                   0.333786   0.002741  121.77   <2e-16 ***
itemY60q:grade                   0.264291   0.001629  162.26   <2e-16 ***
itemY60c:grade:PCMcategorycat_3 -0.121416   0.001776  -68.38   <2e-16 ***
itemY60d:grade:PCMcategorycat_3 -0.154930   0.001808  -85.70   <2e-16 ***
itemY60e:grade:PCMcategorycat_3 -0.167176   0.001422 -117.60   <2e-16 ***
itemY60i:grade:PCMcategorycat_3 -0.214523   0.001791 -119.79   <2e-16 ***
itemY60j:grade:PCMcategorycat_3 -0.155583   0.001753  -88.76   <2e-16 ***
itemY60k:grade:PCMcategorycat_3 -0.185381   0.001836 -100.97   <2e-16 ***
itemY60m:grade:PCMcategorycat_3 -0.092394   0.002969  -31.12   <2e-16 ***
itemY60q:grade:PCMcategorycat_3 -0.131522   0.001817  -72.40   <2e-16 ***
itemY60c:grade:PCMcategorycat_4 -0.173471   0.001713 -101.27   <2e-16 ***
itemY60d:grade:PCMcategorycat_4 -0.322207   0.001794 -179.62   <2e-16 ***
itemY60e:grade:PCMcategorycat_4 -0.380471   0.001531 -248.44   <2e-16 ***
itemY60i:grade:PCMcategorycat_4 -0.412831   0.001825 -226.23   <2e-16 ***
itemY60j:grade:PCMcategorycat_4 -0.247074   0.001718 -143.85   <2e-16 ***
itemY60k:grade:PCMcategorycat_4 -0.369488   0.001836 -201.29   <2e-16 ***
itemY60m:grade:PCMcategorycat_4 -0.277039   0.002819  -98.26   <2e-16 ***
itemY60q:grade:PCMcategorycat_4 -0.292111   0.001780 -164.09   <2e-16 ***
---
```

Appendix D

Explanatory Unconditional Partial Credit Model of the 6-item Measure of Empowerment

```
> summary(model.epcm4)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: PCM ~ -1 + item + item:PCMcategory + (1 | studentnumber)
   Data: empower3
Control: control

      AIC       BIC    logLik  deviance  df.resid
1644225.7 1644457.2 -822093.9 1644187.7   1444178

Scaled residuals:
     Min       1Q   Median       3Q      Max
-12.5206  -0.6962   0.3414   0.5890   5.3531

Random effects:
 Groups        Name        Variance Std.Dev.
 studentnumber (Intercept) 1.883    1.372
Number of obs: 1444197, groups:  studentnumber, 167309

Fixed effects:
                            Estimate Std. Error z value Pr(>|z|)
itemY22b                   3.0995426  0.0263141  117.79   <2e-16 ***
itemY22c                   3.0620951  0.0306936   99.76   <2e-16 ***
itemY22d                   3.3822286  0.0408620   82.77   <2e-16 ***
itemY60l                   2.4196189  0.0121279  199.51   <2e-16 ***
itemY60o                   2.6146930  0.0136643  191.35   <2e-16 ***
itemY60p                   2.9144216  0.0152417  191.21   <2e-16 ***
itemY22b:PCMcategorycat_3 -0.0003911  0.0284150   -0.01    0.989
itemY22c:PCMcategorycat_3  0.4334525  0.0334165   12.97   <2e-16 ***
itemY22d:PCMcategorycat_3  0.6686505  0.0448818   14.90   <2e-16 ***
itemY60l:PCMcategorycat_3 -1.4874480  0.0135832 -109.51   <2e-16 ***
itemY60o:PCMcategorycat_3 -1.2901058  0.0151668  -85.06   <2e-16 ***
itemY60p:PCMcategorycat_3 -1.4137879  0.0166221  -85.05   <2e-16 ***
itemY22b:PCMcategorycat_4 -2.8778156  0.0268477 -107.19   <2e-16 ***
itemY22c:PCMcategorycat_4 -2.3980974  0.0311596  -76.96   <2e-16 ***
itemY22d:PCMcategorycat_4 -1.9944144  0.0412706  -48.33   <2e-16 ***
itemY60l:PCMcategorycat_4 -3.1459568  0.0137945 -228.06   <2e-16 ***
itemY60o:PCMcategorycat_4 -2.9674362  0.0149858 -198.02   <2e-16 ***
itemY60p:PCMcategorycat_4 -3.3216248  0.0164000 -202.54   <2e-16 ***
---
```

Appendix E

Explanatory Main Effects Partial Credit Model of the 6-item Measure of Empowerment

```
> summary(model.epcm5)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: PCM ~ -1 + item + item:PCMcategory + PCMcategory:grade + (1 |
studentnumber)
   Data: empower3
Control: control

     AIC      BIC   logLik deviance df.resid
 1640956  1641224  -820456  1640912  1444175

Scaled residuals:
     Min       1Q   Median       3Q      Max
-12.5721  -0.6928   0.3406   0.5842   5.9707

Random effects:
 Groups        Name        Variance Std.Dev.
 studentnumber (Intercept) 1.855    1.362
Number of obs: 1444197, groups:  studentnumber, 167309

Fixed effects:
                           Estimate Std. Error z value Pr(>|z|)
itemY22b                   2.570878   0.041734   61.60   <2e-16 ***
itemY22c                   2.591739   0.043282   59.88   <2e-16 ***
itemY22d                   2.877345   0.052339   54.98   <2e-16 ***
itemY60l                   1.870369   0.035171   53.18   <2e-16 ***
itemY60o                   2.071409   0.035465   58.41   <2e-16 ***
itemY60p                   2.366046   0.036023   65.68   <2e-16 ***
itemY22b:PCMcategorycat_3  0.487249   0.043265   11.26   <2e-16 ***
itemY22c:PCMcategorycat_3  0.865048   0.045388   19.06   <2e-16 ***
itemY22d:PCMcategorycat_3  1.124296   0.055696   20.19   <2e-16 ***
itemY60l:PCMcategorycat_3 -0.959097   0.035835  -26.76   <2e-16 ***
itemY60o:PCMcategorycat_3 -0.767620   0.036174  -21.22   <2e-16 ***
itemY60p:PCMcategorycat_3 -0.889262   0.036769  -24.18   <2e-16 ***
itemY22b:PCMcategorycat_4 -1.589195   0.041600  -38.20   <2e-16 ***
itemY22c:PCMcategorycat_4 -1.167166   0.043211  -27.01   <2e-16 ***
itemY22d:PCMcategorycat_4 -0.732986   0.052332  -14.01   <2e-16 ***
itemY60l:PCMcategorycat_4 -1.836361   0.035185  -52.19   <2e-16 ***
itemY60o:PCMcategorycat_4 -1.663757   0.035421  -46.97   <2e-16 ***
itemY60p:PCMcategorycat_4 -2.011911   0.035962  -55.95   <2e-16 ***
PCMcategorycat_2:grade     0.062622   0.003981   15.73   <2e-16 ***
PCMcategorycat_3:grade     0.001070   0.002550    0.42    0.675
PCMcategorycat_4:grade    -0.093672   0.002007  -46.68   <2e-16 ***
---
```

Appendix F

Explanatory Full with Interactions Partial Credit Model of the 6-item Measure of Empowerment

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: PCM ~ -1 + item:grade + item:PCMcategory:grade + (1 | studentnumber)
   Data: soccomp3
Control: control

     AIC      BIC   logLik deviance df.resid
 2444500  2444812 -1222225  2444450  1983561

Scaled residuals:
     Min       1Q   Median       3Q      Max
-12.7771  -0.8242   0.4347   0.6992   9.3587

Random effects:
 Groups        Name        Variance Std.Dev.
 studentnumber (Intercept) 1.284    1.133
Number of obs: 1983586, groups:  studentnumber, 162024

Fixed effects:
                                 Estimate Std. Error z value Pr(>|z|)
itemY60c:grade                   0.235557   0.001546  152.38   <2e-16 ***
itemY60d:grade                   0.278059   0.001639  169.69   <2e-16 ***
itemY60e:grade                   0.230899   0.001248  185.00   <2e-16 ***
itemY60i:grade                   0.314495   0.001648  190.80   <2e-16 ***
itemY60j:grade                   0.263118   0.001554  169.28   <2e-16 ***
itemY60k:grade                   0.305438   0.001679  181.87   <2e-16 ***
itemY60m:grade                   0.333786   0.002741  121.77   <2e-16 ***
itemY60q:grade                   0.264291   0.001629  162.26   <2e-16 ***
itemY60c:grade:PCMcategorycat_3 -0.121416   0.001776  -68.38   <2e-16 ***
itemY60d:grade:PCMcategorycat_3 -0.154930   0.001808  -85.70   <2e-16 ***
itemY60e:grade:PCMcategorycat_3 -0.167176   0.001422 -117.60   <2e-16 ***
itemY60i:grade:PCMcategorycat_3 -0.214523   0.001791 -119.79   <2e-16 ***
itemY60j:grade:PCMcategorycat_3 -0.155583   0.001753  -88.76   <2e-16 ***
itemY60k:grade:PCMcategorycat_3 -0.185381   0.001836 -100.97   <2e-16 ***
itemY60m:grade:PCMcategorycat_3 -0.092394   0.002969  -31.12   <2e-16 ***
itemY60q:grade:PCMcategorycat_3 -0.131522   0.001817  -72.40   <2e-16 ***
itemY60c:grade:PCMcategorycat_4 -0.173471   0.001713 -101.27   <2e-16 ***
itemY60d:grade:PCMcategorycat_4 -0.322207   0.001794 -179.62   <2e-16 ***
itemY60e:grade:PCMcategorycat_4 -0.380471   0.001531 -248.44   <2e-16 ***
itemY60i:grade:PCMcategorycat_4 -0.412831   0.001825 -226.23   <2e-16 ***
itemY60j:grade:PCMcategorycat_4 -0.247074   0.001718 -143.85   <2e-16 ***
itemY60k:grade:PCMcategorycat_4 -0.369488   0.001836 -201.29   <2e-16 ***
itemY60m:grade:PCMcategorycat_4 -0.277039   0.002819  -98.26   <2e-16 ***
itemY60q:grade:PCMcategorycat_4 -0.292111   0.001780 -164.09   <2e-16 ***
---
```